ISSN 1870-4069

Analysis of Public Databases of the Health Sector for Decision Making in Health Infrastructures through Artificial Intelligence

Agustín Grajales Castillo[1], Arieh Roldan Mercado Sesma[2], Virgilio Zúñiga Grajeda[3], Felipe Orozco Luna[4], Luis A. Medellín Serna[5], Raúl C. Baptista Rosas[2]

[1] Universidad de Guadalajara,
Centro Universitario de Tonalá,
Programa de Investigación Multidisciplinaria en Salud,
Mexico

[2] Universidad de Guadalajara,
Centro Universitario de Tonalá,
Departamento de Salud,
Mexico

[3] Universidad de Guadalajara,
Centro Universitario de Tonalá,
Departamento de Ciencias Computacionales,
Mexico

[4] Universidad de Guadalajara,
Centro de Análisis de Datos y Supercomputación,
Mexico

[5] Universidad de Guadalajara,
Centro Universitario de Ciencias Exactas e Ingeniería
Mexico

agustin.grajales3462@alumnos.udg.mx, forozco@cads.udg.mx,
{arieh.mercado, virgilio.zuñiga, luis.medellin,
raul.baptista}@academicos.udg.mx

Abstract: One of the priorities that requires urgent responses in Mexico is health. Currently, decisions to build hospitals and health centers are general and centralized. However, they should be based on existing evidence in the information available from public health databases such as distances between municipalities, number and levels of hospitals, poverty and mortality levels, to name a few, and not only on political decisions. This research focused on analyzing with artificial intelligence methods, (decision trees, random forests and bagging technique) of three disaggregated five-year periods, and obtaining a new level of knowledge based on the interpretation of data, for effective decision-making in health infrastructures in the state of Jalisco, Mexico. Decision trees and random forests give us high precision to disperse health infrastructures favoring more inhabitants. The statistical analysis carried out guides a change in the allocation of health systems. The results of this work demonstrate the need to direct/distribute health centers based on these findings.

Keywords: Artificial intelligence; machine learning; big data; decision making; health infrastructures.

1 Introduction

The national health situation in Mexico is extremely delicate. The hospital infrastructure is overwhelmed by the demand for services. While access to health services is restricted to the eligible population: 45% of households in Mexico, the rest, the population with economic resources or informal capacity, are served in private health schemes (out-of-pocket expenses), or they are not taken care of [1].

Geography also helps to complicate the problem, in the main cities, where the specialized infrastructure is located, it is not very accessible for a large part of the population, thus, to the complicated problem of accessing specialized services, distance is added, which represents critical and valuable time.

This directly affects the greater need for care in patients of different age groups and high degrees of specialization that currently do not exist, plus a budgetary burden that announced cuts to the health system [1, 2].

Currently, the official formats for the creation of health infrastructures go through the reading of procedures and filling in imprecise or biased formats [3, 4], which contributes to creating hospitals of a certain level in regions where it is not required or far from a large part of the population.

In the world, Korea and Japan have the highest rate of beds per 1,000 inhabitants, above 12. Followed by some economies in Europe with more than 4, United States 2.8; and the best positioned in Latin America is: Chile (2.01) Colombia (1.69) and Costa Rica (1.15) while in Mexico it is 0.9 [5, 6]. At the national level, Mexico City is the one with the highest availability of beds (2.4) followed by Nuevo León (1.3) and Jalisco (1.2).

Access to healthcare is a requirement for human well-being that is constrained, in part, by the allocation of healthcare resources relative to the geographically dispersed human population [7]. Furthermore, inequities in access to health care contribute to persisting disparities in health care outcomes [8].

It is well known that not all the population has access to a car, so they must reach a health center by public transport, which leads to another factor added to the problem of accessibility: a greater impact on health per se; due to circumstances attributable to traffic given that public transport infrastructure affects road traffic volumes and influences the choice of transport mode [9].

Although we are far from having universal access to health care, which would demand availability and accessibility of services for those who most need health services 10, the contribution of this research is relevant in the search to bring as much as possible, health infrastructures in the most remote or dispersed areas.

In China, Lan et al. published a study on the implications of building hospital infrastructure based on the needs and equipment that serve as a fundamental axis, that is, as a stage prior to the rest, with which the decision of the construction of the physical infrastructure per se, would be carried out if it is included in the hospital infrastructure promotion plans [11].

Table 1. Health establishments in Jalisco.

Type	First level	Second level	Third level	Others[1]
Quantity	1,492	245	16	81

These decisions, as in Mexico, are usually centralized, which also leaves a dilemma: build from the needs but that are not in the budget or, build infrastructures and do not have the budget to equip them. Concepts such as equity and efficiency are sometimes not compatible in health settings. The use of *artificial intelligence*, enabled the use of large amounts of data [12, 13, 14] in all sectors. In medicine, it marks the beginning of a strong impact in three aspects:

1 in clinics, the rapid interpretation of images with high degrees of accuracy;
2 in healthcare systems [15], improving workflow and the potential to significantly reduce medical (human) errors; and;
3 in patients, to enable them in the self-monitoring process for health promotion [16].

This research evaluates the information available in open access databases in the health sector, with artificial intelligence methods to generate predictive models and solve health infrastructure problems, with better evidence-based decision-making and trigger key access points to the health infrastructure, and provide coverage to the currently dispersed or unprotected population of these services.

2 Health Infrastructure in Jalisco

Jalisco is one of the most important economies in Mexico. Its capital: Guadalajara is the third most populated after Mexico City and Mexico state, respectively; is considered the city with the greatest potential for attracting investment in Mexico and second in economic potential in North America and together with five other entities share the largest *GDP* (Internal Product Gross).

Jalisco has 125 municipalities, an educational platform that integrates more than 140 higher level institutions, among which the nearly 50 campuses of the main 30 Universities stand out; the second oldest and largest University in the nation; 6 private universities of national dimension and strong international projection; 2 Advanced Research Centers, as well as one of the largest and most integrated networks of technological education institutions in Mexico [17].

Guadalajara together with nine other cities around it form the metropolitan area that cluster the most important *infrastructure of health* from public to private hospitals (supplemental material S1).

Due to the population of the eastern state of Jalisco must move to other places where the infrastructure and specialized services are found (second and third level) resulting in longer travel times, man-hours of no labor production, and above all, demand for specialized personnel.

[1] Warehouses, laboratories, vaccination centers, administrative offices

Table 2. Third level health establishments in Jalisco.

Municipalities	**Guadalajara**	**Zapopan**	**Zapotlanejo**	**Tlajomulco**	**Tala**	**San Miguel**	**Ocotlán**	**Total**
Third level	6	5	1	1	1	1	1	16

Table 3. Second level health establishments in Jalisco.

Municipalities	Guadalajara	Zapopan	Arandas	Zapotlán	Puerto Vallarta	Tepatitlán	Tonalá	Tlaquepaque	Ocotlán	Tlajomulco	Total
Second level	88	26	9	9	9	7	5	5	5	4	167

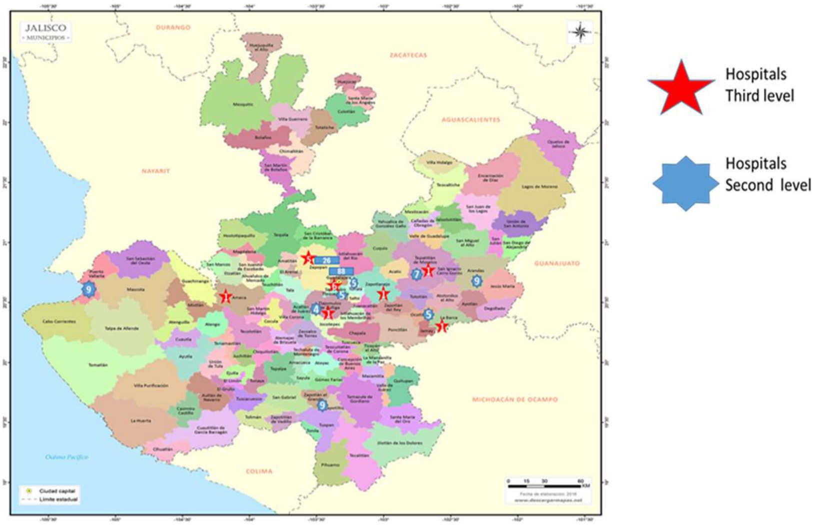


Fig 1. Municipalities of Jalisco. Adapted from: municipalities of Jalisco map with names - Google Search.

The state of Jalisco has 1,834 Health establishments, called CLUES (Unique Code of Health Establishments, IIEG) [18], classified as presented in table 1. While 511 are classified as Private Medical Services.

Tables 2 and 3 show the distribution of health infrastructures in the state; 7 municipalities concentrate 100% of level 3 and 10 municipalities 68% of level 2 hospitals. These levels of care are the basis of the National Health System. The concentration of these hospitals is in the metropolitan area of Guadalajara (figure 1).

From this perspective, it is essential to propose more comprehensive solutions, which go beyond installing orientation modules, accessibility and updated quasi-electronic files, among other things. These decisions must be supported by evidence considering the main variables: geography, distance between municipalities, number of hospitals, mortality levels and poverty levels, among others, and not only in political decisions.

However, the secretaries or government agents who decide on the growth of the infrastructure use few of these variables to resolve the creation of second and third level health centers or hospitals [27, 28], due to the large amount of data that needs to be processed. They simply do not analyze data.

Fig 2. Workflow diagram.

Table 4. Selection and description of variables.

Variable	**Description**
Poblacion	population in the five-year period
Pobreza	population in poverty in the five-year period
Defunciones	deaths in the five-year period
TasaMortalidad	deaths / population in the five-year period
PoblHosp	population / sumHosp1,2 in the five-year period
DefvsPob	deaths / poverty in the five-year period
DensidadxHabitante	population / surface km2
Medicos100kJal	260 doctors / 100, 000 inhabitants *[2]
NumHosp	number of 2nd and 3rd level hospitals
DistMun	90 km intermunicipal distance
NumMun	municipality number
Zone	zone
Mun	municipality name

This research focuses on how *AI* and *ML* [19] technological tools, can contribute to the analysis, design and modeling of data, in the municipalities of the state of Jalisco, to diagnose optimal solutions generation of health infrastructures; with the available information contained in free access databases related to the health sector.

The model intends to detonate key points, for the solution of new access points to the health infrastructure and cover the population currently dispersed or unprotected from these services.

3 Methods

3.1 Software Used

The programming language R version 4.1.3[3] [20] and the RStudio platform version 4.1.1[4] were used [21], which is a language with a statistical approach to data science.

3.2 Workflow Diagram

The workflow diagram is presented in figure 2.

[2] Taken from [27]
[3] https://cran.r-project.org/
[4] https://www.rstudio.com/

Table 5. Some variables do not provide weighty information to the model.

sex	**age**	**scolarship**	**occupation**	**birthday**	**day of certificacion**
necropsy	marital status	successor	preagnancy	birth month	month of certification

3.3 Data Transformation

Data mining [22] through data filtering, analysis and visualizations was applied in the work files to transform the information and make it readable by the software (figure 2).

Quantitative variables of population, poverty, deaths, mortality rate, population by hospital, deaths by population, per capita density, doctors per 100,000 inhabitants in Jalisco, number of hospitals, distances between municipalities, number of municipalities, zone and name of municipalities, of three disaggregated five-year periods (2010, 2015, 2020 INEGI) [23, 24, 25] were considered; applying statistics and machine learning classification algorithms: decision trees, random forests and bagging technique, to contrast results.

Although there are various techniques for classification, *decision trees* and *random forest* due to their short training time and results with high values of precision in classification are obtained. While other methods such as *support vector machines (SVM)* technique to classify has in the background the idea of finding the best hyperplane by which to separate the data. They work quite well for text classification, for example.

3.4 Variables Selection and Feature Extraction

Were include (13/50) variables [24, 25] (table 4), that they would contribute to the development of the model; and exclude others, with this, eliminate the “malediction of multidimensionality” [26] (reduction of redundant, spurious variables or those that do not provide weighty information to the model, see table 5). In this way, the relationships between the variables are better understood.

3.5 Processing

Were grouped municipalities in the variables NumMun (number of municipality), Zone, Mun (name of municipality) that were not more than 90 km apart (DistMun); (the above as a guideline that each access to a health infrastructure be 90 minutes away, considering a speed of 60 km/hour) and the number of second and third level hospitals (NumHosp).

The rest of variables were categorized into low, medium, high levels; to identify the different possibilities of variation. *Decision trees* and *random forests* are generated, and also the *bagging* technique, data partition and prediction (supplemental material S2), of the three five-year periods disaggregated with different variables and the statistical results are reflected.

Table 6. Results with the three algorithms of the three five-year periods.

Algorithm	Variables	Period	Sensitivity %	Specificity %	Accuracy %
Decision trees	10[5]	five-year period 2010-2014	100	100	100
Random forest	8[6]	five-year period 2015-2019	100	75	88
Decision trees	9[7]	five-year period 2020	100	100	100
Bagging	9[5]	five-year period 2010-2014	100	75	88

4 Results

The results are presented with the three algorithms of the three five-year periods with 10, 8 and 9 variables respectively, see table 6. We used the metrics of Sensitivity, Specificity (correct/incorrect classification) and balanced Accuracy.

For the five-year period 2010-2014, with 10 variables, the best result is presented in the *decision trees* at 100%: Zapotiltic (table 7 and map 1 supplemental material S3) would benefit 23 municipalities in 4 zones, covering a population of 439,904 inhabitants.

For the five-year period 2015-2019, with 8 variables, the best result is presented in *random forest* at 88%: San Miguel El Alto (table 8 and map 2 supplemental material S3) would benefit 19 municipalities in 3 zones, covering a population of 787,740 inhabitants. For the five-year period 2020, with 9 variables, the best result is presented in *decision trees* at 100%: La Barca (table 9 and map 3 supplemental material S3) would benefit 15 municipalities in 3 zones, covering a population of 960,253 inhabitants.

With ten variables, the models behave more stable, except for *bagging* (multiclassifier), which is stricter. With eight variables, the best results were in *random forests* at 88% and with nine variables the best results were in *decision trees* at 100% while *bagging* technique reached at 88% as highest result.

Population, poverty and deaths are recorded in all the models as they are the center of the analysis. Categorizing variables offers an important contribution (Figures A1 and A3, supplemental material S4). Mortality rate and deaths by population had the highest weight of in the models.

There is homogeneity in the variables. The different percentages of *Accuracy* in the results concentrate guide to maintain an objective of 80% in the different algorithms to consider the model stable, robust and with the possibility of improvement by incorporating and categorizing more weight variables.

---

[5] Number of municipalities, zone, name of municipalities, distances between municipalities, poverty, deaths, population, mortality rate, population by hospital, deaths by population

[6] Number of municipalities, zone, name of municipalities, distances between Municipalities, poverty, deaths, population, mortality rate

[7] Number of municipalities, zone, name of municipalities, distances between municipalities, poverty, deaths, population, per capita density, doctors per 100k inhabitants in Jalisco

A classification system is considered useful when it has a higher accuracy rate in the majority class (positives vs. negatives). The current processes to detonate *health infrastructures* in Mexico do not consider statistical methods, so this percentage is considered highly significant for decision making.

Supplementary material S3 shows the tables with the zones, municipalities, distances between municipalities, the coverage population of the municipalities, and corresponding coverage maps after the analysis. In addition, the map of the state of Jalisco with the three models after the analysis.

5 Conclusions

In the explored databases, there is a lot of useful information for the purposes of this research, there are variables that highlight the importance of considering the distances between the municipalities, the number of hospitals, deaths and poverty levels; it is well known that the higher the level of poverty, the greater the marginalization and, consequently, deaths. It is clarified that there are some other data that do not represent the leverage for the development of the model, as explained in tables 4 and 5; however, these latest data could be useful for health decision makers to better equip hospitals according to the type of disease, for example.

The national health situation in Mexico requires concrete and reliable responses in health systems. People's lives are involved. They must be based on a base of concrete, logical information and, where required, forecast and/or prevention. It is in these scenarios where *decision-making* 27 has become an “art”, and they cannot occur without first comprehensively analyzing the information available in the Health databases. In the work carried out to date, the advances presented suggest that, with an optimal combination of variables, and statistical and algorithmic *data analysis*, the resulting information is much more objective and precise.

It is evident that forecasting an infrastructure and not installing it (which would have saved lives) is not the same as installing it when it is not efficient in terms of serving the inhabitants, distances between municipalities, among other metrics. The results of the work demonstrate the need to direct/distribute the centers based on these findings.

As seen in the tables and maps generated in supplemental material S4, it is essential to cover the eastern areas of the state of Jalisco, since the actual distribution is concentrated in the metropolitan area, leaving these municipalities out and contributing to social inequality and economic backwardness. With this, the number of available beds in the state would rise, since, it is currently 1.2 [6].

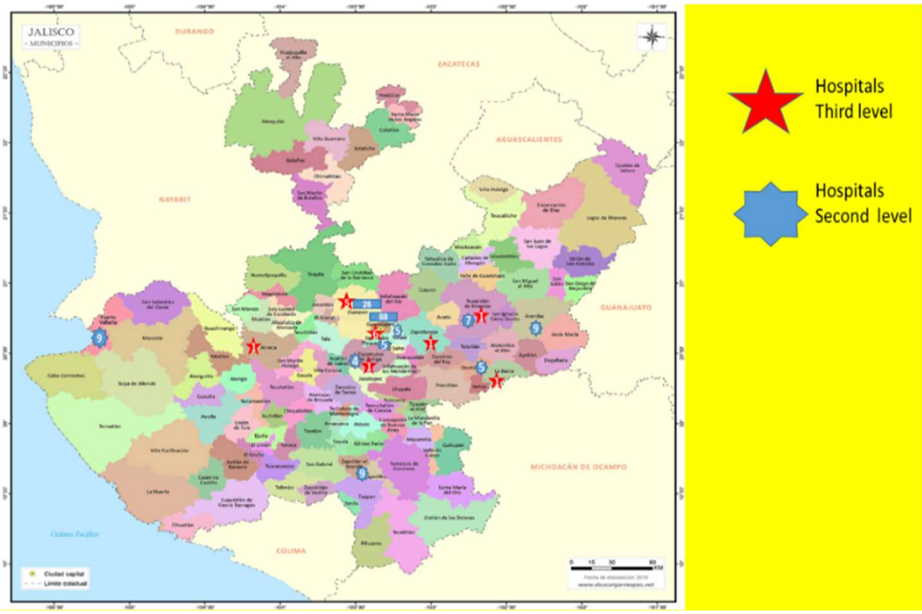
In this multidisciplinary line of research related to the combination of *machine learning* models and *data mining* [22, 29, 30, 31], it is aimed at solving problems and decisions in *health infrastructures*, with a statistical methodology assisted by the *AI*.

The social and scientific value that the research represents is important, since it is aimed at improving the living conditions and well-being of the population of the state of Jalisco; while producing knowledge that opens opportunities for *decision making* and problem solving in health service infrastructures.

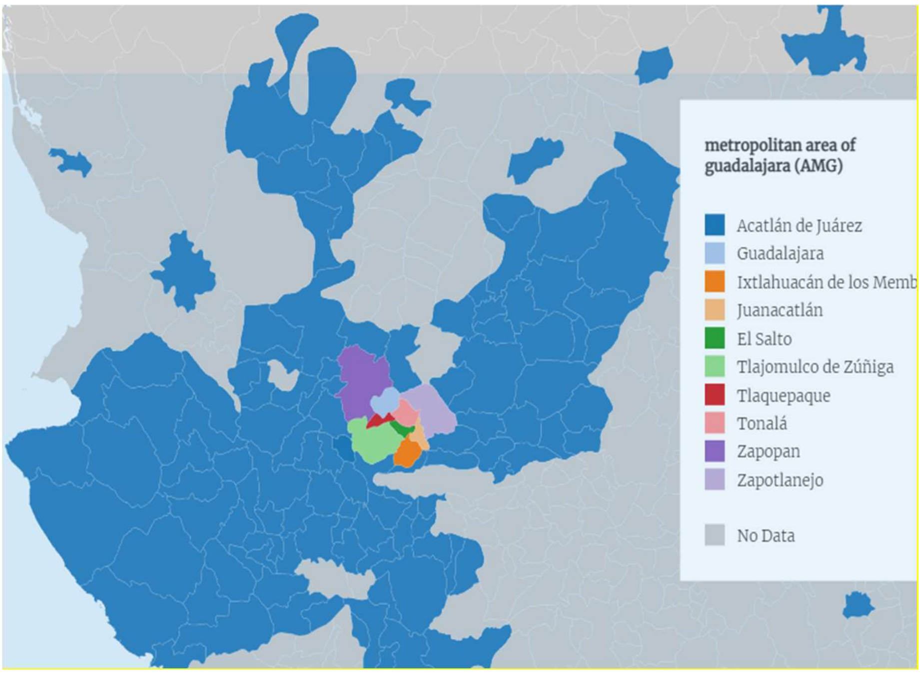
Supplementary materials are presented in the following sections: S1, S2, S3, S4.

Supplementary Material S1

(1) Distribution map of health infrastructures in Jalisco, second and third level and (2) Metropolitan area of Guadalajara (MAG) map.



(2) Figure adapted of: municipalities of Jalisco map with names - Google Search.



Supplementary Material S2

1) Pseudocode for grouping of municipalities and variable categorization.

```
temp <- dataset_distancias[dataset_distancias[,i] <= 90,]
colnames(temp)[i] <- "Distancia"
temp <- temp[,c(3,i)]
temp <- merge(temp,dataset, by=c("Municipio","Municipio"))

temp$factor_Pob220 <- discretize(temp$Poblacion2020,breaks=2,labels=c("Bajo","Alto"))
temp[,3] <- NULL
temp$factor_Pobre220 <- discretize(temp$Personas.en.situacion.de.pobreza2020..,breaks=2,labels=c("Bajo","Alto"))
temp[,3] <- NULL
temp$factor_Def220 <- discretize(temp$Defunciones2020..,breaks=2,labels=c("Bajo","Alto"))
temp[,3] <- NULL
temp$factor_TasaM <- discretize(temp$TASAS.DE.MORTALIDAD.x.100.000,breaks=2,labels=c("Bajo","Alto"))
temp[,3] <- NULL
temp$factor_PobHosp220 <- discretize(temp$PobHospQ2020,breaks=2,labels=c("Bajo","Alto"))
temp[,5] <- NULL
temp$factor_DefvsPobQ2020 <- discretize(temp$DefvsPobQ2020,breaks=2,labels=c("Bajo","Alto"))
temp[,5] <- NULL
```

2) Pseudocode for the corresponding algorithms (2.1 *decision trees*, 2.2 *random forest* and 2.3 *bagging*).

2.1)

```
# trees decision

data <- data %>%
    mutate_at("random_nuevHosp_f", factor)

split = sample.split(data$random_nuevHosp_f, SplitRatio = 2/3)

data_entrenamiento = subset(data, split == TRUE)
data_prueba = subset(data, split == FALSE)

arbol <- rpart(formula = random_nuevHosp_f ~ .,
        data = data_entrenamiento,
                minsplit = 2,
            method = "class")
```

2.2)

```
# random forest

data <- data %>%
    mutate_at("random_nuevHosp_f", factor)

split = sample.split(data$random_nuevHosp_f, SplitRatio = 2/3)

data_entrenamiento = subset(data, split == TRUE)
data_prueba = subset(data, split == FALSE)

arbol <- randomForest(formula = random_nuevHosp_f ~ .,
        data = data_entrenamiento)

# prediction and confusion matrix

pred_arbol <- predict(arbol, newdata = data_prueba, type = 'class')
data_prueba <- cbind(data_prueba,pred_arbol)

cm <- confusionMatrix(pred_arbol, data_prueba[["random_nuevHosp_f"]])
tocsv <- data.frame(cbind(t(cm$overall),t(cm$byClass)))
tocsv_percent <- tocsv
```

2.3)

```
# bagging technique

data <- data %>%
      mutate_at("random_nuevHosp_f", factor)

split = sample.split(data$random_nuevHosp_f, SplitRatio = 2/3)

data_entrenamiento = subset(data, split == TRUE)
data_prueba = subset(data, split == FALSE)

arbol <- bagging(formula = random_nuevHosp_f ~ .,
data = data_entrenamiento)
```

Supplementary Material S3

Tables 1-3 and the corresponding coverage maps (1-3) after the analysis and (4) general map after the analysis.

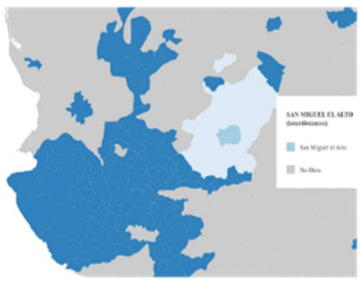
Table 1. Five-year period 2010-2014 after analysis.

NumMun	Zone	Municipalities	Zapotiltic (interdistances)	Population 2020
14004	Lagunas	Amacueca	58	5,770
14014	Lagunas	Atoyac	53	8,730
14023	Sur	Ciudad Guzmán	14	116,094
14057	Sureste	La Manzanilla De La Paz	85	4,118
14059	Sureste	Mazamitla	70	14,629
14065	Sur	Pihuamo	58	11,595
14069	Sureste	Quitupan	90	7,770
14079	Sur	San Sebastián Del Sur	33	17,932
14082	Lagunas	Sayula	43	37,902
14085	Sur	Tamazula De Gordiano	22	36,253
14086	Lagunas	Tapalpa	86	20,770
14087	Sur	Tecalitlán	32	17,365
14089	Lagunas	Techaluta De Montenegro	59	4,091
14092	Lagunas	Teocuitatlán De Corona	79	10,708
14099	Sur	Tolimán	75	11,928
14102	Sierra de Amula	Tonaya	80	5,989
14103	Sur	Tonila	37	7,600
14106	Sierra de Amula	Tuxcacuesco	80	5,508
14108	Sur	Tuxpan	12	38,573
14112	Sureste	Valle De Juárez	79	6,180
14113	Sur	San Gabriel	65	15,933
14119	Lagunas	Zacoalco De Torres	80	26,965
14122	Sur	Zapotitlán De Vadillo	72	7,501
			23	439,904

Map 1. Coverage map five-year period 2010-2014 after analysis

Table 2. Five-year period 2015-2019 after analysis.

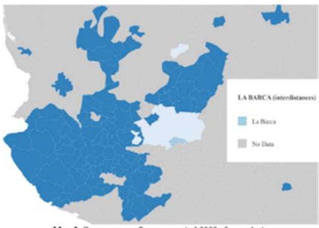
NumMun	Zone	Municipalities	Sn Miguel El Alto (interdistancias)	Population 2020
14001	Altos Sur	Acatic	72	20,644
14008	Altos Sur	Arandas	57	79,688
14013	Ciénega	Atotonilco El Alto	69	56,994
14035	Altos Norte	Encarnación De Díaz	68	48,673
14046	Altos Sur	Jalostotitlán	19	30,917
14048	Altos Sur	Jesús María	78	18,227
14053	Altos Norte	Lagos De Moreno	78	177,818
14060	Altos Sur	Mexticacán	63	5,332
14072	Altos Norte	San Diego De Alejandría	46	7,645
14073	Altos Norte	San Juan De Los Lagos	35	70,418
14074	Altos Sur	San Julián	25	15,284
14091	Altos Norte	Teocaltiche	63	34,545
14093	Altos Sur	Tepatitlán De Morelos	50	133,350
14105	Ciénega	Tototlán	89	20,846
14109	Altos Norte	Unión De San Antonio	60	19,914
14111	Altos Sur	Valle De Guadalupe	27	6,658
14117	Altos Sur	Cañadas De Obregón	40	4,408
14118	Altos Sur	Yahualica De González Gallo	84	19,169
14125	Altos Sur	San Ignacio Cerro Gordo	54	[illegible]
			19	787,740



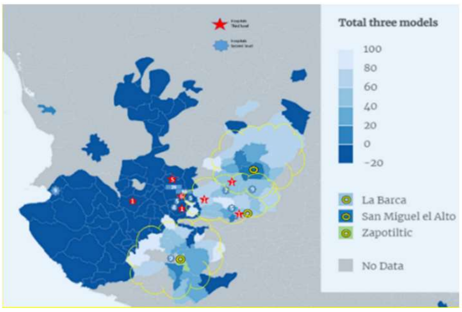
Map 2. Coverage map five-year period 2015-2019 after analysis

Table 3. Five-year period 2020 after analysis.

NumMun	Zone	Municipalities	La Barca (interdistances)	Population 2020
14008	Altos Sur	Arandas	69	79,688
14013	Ciénega	Atotonilco El Alto	34	56,994
14016	Ciénega	Ayotlán	49	42,119
14033	Ciénega	Degollado	70	23,954
14044	Centro	Ixtlahuacán De Los Membrillos	88	65,732
14047	Ciénega	Jamay	18	23,411
14048	Altos Sur	Jesús María	67	18,227
14063	Ciénega	Ocotlán	30	103,173
14066	Ciénega	Poncitlán	47	52,955
14070	Centro	El Salto	87	239,313
14093	Altos Sur	Tepatitlán De Morelos	79	133,350
14105	Ciénega	Tototlán	44	20,846
14123	Ciénega	Zapotlán Del Rey	48	17,697
14124	Centro	Zapotlanejo	76	65,584
14125	Altos Sur	San Ignacio Cerro Gordo	66	17,210
			15	960,253



Map 3. Coverage map five-year period 2020 after analysis



Map 4. Coverage map of the three models generated after analysis and their area of influence; second and third level hospital are represented.

SupplementaryMaterial S4

Variables categorization and contribution in generated models.

Figure A1

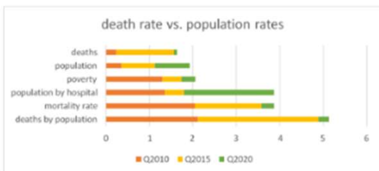


Figure A2

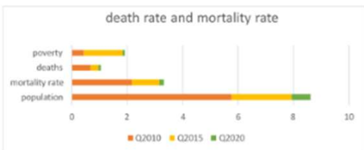
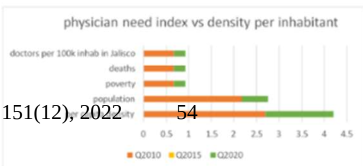


Figure A3



References

1. Jiménez, M.: Sistema de salud pública, un problema más para los mexicanos. Forbes México (2016) https://www.forbes.com.mx/sistema-de-salud-publica-un-problema-mas-para-los- mexicanos/
2. García, A.: Los retos para mejorar el Sistema de salud pública en México. El Economista (2019) https://www.eleconomista.com.mx/estados/Los-retos-para-mejorar-el-sistema-de-salud-publica-en-Mexico-20190105-0001.html
3. Secretaría de Salud: Plan maestro de infraestructura física en salud. Dirección General de Planeación y Desarrollo en Salud (2023) https://www.gob.mx/salud/acciones-y-programas/plan-maestro-de-infraestructura-fisica-en-salud
4. Secretaría de Salud: Certificado de necesidad. Dirección General de Planeación y Desarrollo en Salud (2022) https://www.gob.mx/salud/acciones-y-programas/certificado-de-necesidad?state=published
5. Statista: Hospital bed density select countries 2020. Health Proffesionals and Hospitals (2020) https://www.statista.com/statistics/283273/oecd-countries-hospital-bed-density/
6. Secretaria de Salud: Sistema de Información de la Secretaría de Salud (2018) http://sinaiscap.salud.gob.mx:8080/DGIS/
7. Weiss, D. J., Nelson, A., Vargas-Ruiz, C. A., Glicoric, K., Bavadekar, S., Grabilovich, E., Bertozzi-Villa, A., Roizer, J., Gibson, H. S., Shekel, T., Kamath, C., Lieber, A., Schulman, K., Shao, Y., Qarkaxhija, V., Nandi, A. K., Keddie, S. H., Rumisha, S., Amratia, P., Arambepola, R., Chestnutt, E. G.: Global maps of travel time to healthcare facilities. Nature Medicine, vol. 26, pp. 1835–1838 (2020) doi: 10.1038/s41591-020-1059-1
8. Guo, J., Hernandez, I., Dickson, S., Tang, S., Essien, U. R., Mair, C., Berenbrok, L. A.: Income disparities in driving distance to health care infrastructure in the United States: A geographic information systems analysis. BMC Research Notes, vol. 15, no. 225 (2022) doi: 10.1186/s13104-022-06117-w
9. Tétreault, L. F., Eluru, N., Hatzopoulou, M., Morency, P., Plante, C., Morency, C., Reynaud, F., Shekarrizfard, M., Shamsunnahar, Y., Faghih-Imani, A., Drouin, L., Pelletier, A., Goudreau, S., Tessier, F., Gauvin, L., Smargiassi, A.: Estimating the health benefits of planned public transit investments in Montreal. Environmental Research, vol. 160, pp. 412–419 (2018) doi: 10.1016/j.envres.2017.10.025
10. Escamilla, V., Calhoun, L., Winston, J., Speizer, I. S.: The role of distance and quality on facility selection for maternal and child health services in urban Kenya. Journal of Urban Health, vol. 95, no. 1, pp. 1–12 (2018) doi: 10.1007/s11524-017-0212-8
11. Lan, T., Chen, T., Hu, Y., Yang, Y., Pan, J.: Governmental investments in hospital infrastructure among regions and its efficiency in China: An assessment of building construction. Frontiers in Public Health, vol. 9, no. 719839 (2021). doi: 10.3389/fpubh.2021.719839
12. García-Alsina, M.: Big data: Gestión y explotación de grandes volúmenes de datos. Editorial UOC (2017) http://digital.casalini.it/9788491167112
13. López-Murphy, J. J., Zarza, G.: La ingeniería del big data: Cómo trabajar con datos. Editorial UOC (2017)
14. Bonsón Ponte, E., Sierra Molina, G.: Intelligent Accounting: impact of Artificial Intelligence on accounting research and accounting information. In: Proceedings of the ITHURS, León, pp. 361–368 (1996)
15. Hoffman S.: The great promise of artificial intelligence for public health. Research and Innovation (2018) https://yfile.news.yorku.ca/2018/04/05/the-great-promise-of-artificial-intelligence-for-public-health/
16. Topol, E. J.: High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, vol. 25, no. 1, pp. 44–56 (2019) doi: 10.1038/s41591-018-0300-7

17. Instituto de información y estadística geográfica de Jalisco (IIEG) (2022) https://iieg.gob.mx/ns/
18. Sebastian, Y., Tiong, X. T., Raman, V., Fong, A., Then, P.: Advances in diabetes analytics from clinical and machine learning perspectives. International Journal of Design, Analysis and Tools for Integrated Circuits and Systems, vol. 6, no. 1, pp. 32–37 (2017)
19. CRAN: The comprehensive R archive network (2023) https://cran.r-project.org/
20. Posit: The open-source data science company (2023) https://www.rstudio.com/
21. Febles-Rodríguez, J. P., González-Pérez, A.: Aplicación de la minería de datos en la bioinformática. Acimed, vol. 10, no. 2, pp. 69–76 (2002)
22. Instituto nacional de estadística y geografía (INEGI) (2022) https://www.inegi.org.mx/datos/
23. Dirección general de información en salud: Defunciones: Datos abiertos. Secretaría de Salud (2022) http://www.dgis.salud.gob.mx/contenidos/basesdedatos/da_defunciones_gobmx.html
24. Consejo nacional de la política de desarrollo social (CONEVAL): Indicadores de pobreza municipal 2010-2015. Datos Abiertos (2019) https://www.coneval.org.mx/Paginas/principal.aspx
25. Beltrán-Pascual, M., Muñoz-Martínez, A., Muñoz-Alamillos, Á.: Redes bayesianas aplicadas a problemas de credit scoring. Una aplicación práctica, Cuadernos de Economía, vol. 37, no. 104, pp. 73–86 (2014) doi: 10.1016/j.cesjef.2013.07.001
26. Vargas, M.: Jalisco con solo 2.6 médicos por mil habitantes. Centro Universitario de Ciencias de la Salud (2022) https://www.cucs.udg.mx/noticias/archivo-de-noticias/jalisco-con-solo-26-m-dicos-por-mil-habitantes
27. Turban, E., Sharda, R., Delen, D., Aronson, J. E., Liang, T. P., King, D.: Decision support and business intelligence systems. Pearson College Division (2005)
28. Mikhaylov, S. J., Esteve, M., Campion, A.: Artificial intelligence for the public sector: Opportunities and challenges of cross-sector collaboration. Philosophical Transactions of the Royal Society A Mathematical, Physical and Engineering Sciences, vol. 376, no. 2128, (2018) doi: 10.1098/rsta.2017.0357
29. Benítez, R., Escudero, G., Kanaan, S., Rodó, D. M.: Inteligencia artificial avanzada. Editorial UOC (2014)
30. Davenport, T., Kalakota, R.: The potential for artificial intelligence in healthcare. Future Healthcare Journal, vol. 6, no. 2, pp. 94–98 (2019) doi: 10.7861/futurehosp.6-2-94